2008

Expectation-Maximization Optical Tomosynthetic Volume Imaging

Philip M. Hanna
Wright State University

EXPECTATION-MAXIMIZATION OPTICAL TOMOSYNTHETIC VOLUME IMAGING

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

by

PHILIP M. HANNA
Department of Electrical Engineering
Wright State University

2008
Wright State University

WRIGHT STATE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

June 05, 2008

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Philip M. Hanna ENTITLED Expectation-Maximization Optical Tomosynthetic Volume Imaging BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosphy.

Brian D. Rigling, Ph.D.
Dissertation Director

Ramana V. Grandhi, Ph.D.
Director, Ph.D. in Engineering Program

Joseph F. Thomas, Jr., Ph.D.
Dean, School of Graduate Studies

Committee on
Final Examination

Brian D. Rigling, Ph.D.

Fred D. Garber, Ph.D.

Lang Hong, Ph.D.

Randolph L. Moses, Ph.D.

Michael L. Bryant, Ph.D.

ABSTRACT

Hanna, Philip M. Ph.D. Egr., Department of Electrical Engineering, 2008. Expectation-Maximization Optical Tomosynthetic Volume Imaging.

Optical tomosynthetic imaging of 3–D objects from disparate 2–D images has been limited in the past by a lack of algorithmic enforcement of physical constraints, in particular within-scene obscuration and object self-occlusion. This paper presents a stochastic observation model of a tomosynthetic collection that explicitly includes an obscuration operator that is unknown by the sensor. The expectation-maximization algorithm is used to iteratively estimate the obscuration operator and to reconstruct the 3–D volume of interest. Explicit inclusion of obscuration effects greatly enhances the spatial and spectral accuracy of 3–D results without use of costly post-processing techniques. Performance metrics are introduced and the resultant receiver operating characteristics are presented.

CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENT

I would first like to think my advisor, Dr. Brian Rigling, for all of his help, direction and patience. I would also like to thank the Air Force Research Laboratory's Sensors Directorate, Automatic Target Recognition Division for supporting this research. Finally, I would like to thank my family for all of their understanding, sacrifice and support.

CHAPTER 1
INTRODUCTION

It is the responsibility of the assisted target recognition (ATR) community to develop assets needed to accurately discriminate threats. In order to do this we would like to be able to develop and deploy a low-cost, persistent-surveillance asset to capture high-resolution, three-dimensional target and threat environment data for use in ATR system development. One potential asset could be an unmanned airborne vehicle (UAV) carrying a video imager. However, passive electro-optic imaging systems are presently limited by their ability to provide only 2–D observations. It is thus our goal to analyze and enhance the employment of this sensor platform to obtain a 3–D estimate of a target in a scene of interest based on the airborne observations made by an optical 2–D sensor array (i.e., video imager). Ideally, a high-fidelity 3–D reconstruction is achieved with only a single pass of the UAV through the scene environment.

The development of a 2–D to 3–D algorithm that takes the imager output and reconstructs the scene is of paramount importance to the ATR community. The tomosynthetic technique presented here shows the potential capability to utilize existing surveillance assets and process the imager observations to reconstruct a scene based on the physical characteristics of the target. There are a number of applications that are expected to benefit from the proposed method. First (and primarily), this technique will extract 3–D data from a single aperture sensor performing a single pass taking only 2-D measurements. Second, obtaining 3–D data from 2-D measurements provides the ATR community with a capability to populate signature databases with-

out the costly measurement techniques currently in use. Third, this method shows great noise and clutter rejection capability. Finally, this technique will, for the most part, be a software only upgrade to existing sensor assets, saving valuable resources.

There has been considerable research invested in 2–D to 3–D reconstruction methods. From the image processing community, we have shape estimation and super-resolution. From the medical imaging community, we have computed tomography (CT) and tomosynthesis. Traditional tomosynthetic processing of 2–D video data can synthesize a scene volume; however, extracting 3–D object components from that scene is costly in terms of computational requirement [6]. In addition, traditional tomosynthesis noise reduction techniques are not applicable to non-cooperative surveillance measurement scenarios, since traditional tomosynthesis noise reduction pertains to the use of the Poisson noise model for the statistical nature of the nuclear medical imaging detector versus the Gaussian noise model that is commonly used to model the entire sensor system and observation environment [10, 15].

Shape estimation algorithms include a large class of computer vision techniques that utilize known geometry to estimate the surface of an object [16]. Some of these include: structure from motion (SFM) and structure from planar motion (SFPM) [7], shape from stereo (SFS) [17], depth from focus (DFF) [13] and depth from defocus (DFD)[14]. Structure from motion and shape from stereo both fall under the class of triangulation techniques. The main difference between these two is that SFS usually involves the simultaneous measurement of two points (hence the term stereo), and SFM can be considered the temporal version of SFS, with the motion of the sensor (or in the case of [7], motion of the object with stationary sensor) at the next interval of time bringing it to the same position as the second measurement site as in the stereo sensor. SFM, SFS, and SFPM are triangulation techniques, where the optics generally are unchanging, and the angular relationship between the scene and the sensor changes during observations. DFF and DFD are similar to each other in that

they estimate the depth by searching for the state of the imaging system for which the object is in (or out of) focus. This is done by changing the optical geometry, either by varying the focal length or object distance (or a combination of both) [14].

The field of tomographic imaging has been growing for nearly eighty years, with the basic principles of tomography formulated in 1932 by Ziedses des Plantes [5]. As technology progressed, this evolved into the common X-ray computed tomography (CT). Then, in 1972, a seminal paper on tomosynthesis was published by David G. Grant[5, 12, 2]. Grant's work presented tomosynthesis as a method to reconstruct an arbitrary cross section of an imaged object by using a 3–D back-projection algorithm. It offered advantages over the more common CT such as lower costs, lower radiation dosage, and simultaneous multi-plane reconstruction. Grant's tomosynthesis was confined to environments of interest to the radiology community of that time, producing three-dimensional X-ray images in a tightly controlled environment where noise was of minimal concern.

When the medical imaging community progressed from X-ray CT and tomosynthesis to emission tomography (ET) measurement noise became a problem [15]. The models for emission tomography did not distinguish the physics from that of transmission tomography. As positron emission tomography (PET) began to become popular, researchers developed general mathematical models for ET based on the physics of PET which utilizes the emission and detection of photons in its imaging mechanism via a Poisson processes [15, 2] . In particular, improvements in image reconstruction from emission computed tomography (ECT) data were achieved by using maximum-likelihood (ML) estimation to estimate the radioactive distribution. Because of the shift-and-add process [5] used in tomosynthesis, back-projection resulted in blurred energy from elements that were not in the focal plane mixing in with the spatially well-defined energy from elements located on the focal plane. This results in a major shortcoming of the backprojection algorithm: artifacts in the reconstructed volume.

Chen and Barner introduced a 3–D ML reconstruction technique that updated the reconstruction via a statistical model of the imaging energy and greatly improved the blurred image artifacts as a result [2, 1].

A number of researchers proposed the use of the expectation-maximization (EM) algorithm to maximize the Poisson likelihood estimates in PET, and achieved positive results [15, 3]. The maximum-likelihood expectation-maximization (ML–EM) has become the de facto state of the art image reconstruction approach in the PET as well as in the single photon emission computed tomography (SPECT) community [6]. There are a number of documented advantages of ML–EM over the conventional filtered back-projection algorithm used in ECT. Examples of this include significant improvements in computational speed [9, 3] and better image quality with fewer artifacts [2].

With an end goal of target signature exploitation, we are interested in methods that take the physical properties of the target into account to enhance the reconstruction. This has led us to an approach that builds off of the tomosynthesis reconstruction method, extending it with an estimator that utilizes information inferred from the target spectral characteristics to provide a high fidelity reconstruction from the observations. Novel application of the expectation-maximization algorithm is used to reduce measurement noise effects on the scene reconstruction and extend the medical imaging community's tomosynthetic techniques by EM estimation of the RGB color characteristics of the target and scene, ensuring the accurate construction of 3–D target elements while de-emphasizing obscurations.

This dissertation extends traditional tomosynthesis through the use of a 3-color airborne video imager, as opposed to a monochromatic non-visible electromagnetic radiation source, as well as implementing the ML–EM algorithm to not only exploit the spectral characteristics of that sensor, but also to iteratively refine an estimate of object self-obscuration. The proposed EM approach will thereby produce a volume

estimate that will not only have a more accurate spatial representation of the true volume, but will also have a more accurate spectral (RGB) color representation. This can be applied directly to higher-dimensioned color spaces (e.g., hyper-spectral imaging).

In particular, we extend the ML-EM algorithm from the Poisson likelihood used in ECT [8] for describing material densities in a tightly controlled medical measurement environment to our model of Gaussian measurement noise from an airborne sensor flying around a scene of interest. We utilize characteristics of our video imaging sensor as the parameter set that the EM algorithm seeks to estimate, instead of the material densities of the sample under examination as was described in [6].

Chapter 2 of this dissertation will present a detailed problem statement, followed by a description of traditional tomosynthesis as it would apply to this scenario (Chapter 3). We extend the traditional tomosynthesis technique to the new EM tomosynthesis reconstruction, and present how the obscuration estimate (Chapter 4) evolves through use of the ML–EM estimation algorithm (Chapter 5). We then present a description of a computer simulation that takes a scene and images it based on a theoretical measurement model. We show how the initial estimate is created using traditional tomosynthesis, and how the ML–EM estimator iterates over successive estimated volumes to arrive at a reconstructed scene (Chapter 6). We present a performance evaluation of that algorithm (Chapter 7). Finally, we discuss our thoughts on future directions for this approach.

CHAPTER 2
PROBLEM STATEMENT

We seek to obtain a 3–D estimate of a scene of interest based on airborne observations by an optical 2–D sensor array (i.e., camera). Figure 2.1 illustrates the sensor/measurement geometry for the observation scenario of interest. In this measurement environment, an airborne sensor platform captures a sequence of geo-referenced, electro-optical images. Here, we assume that the images are obtained from a standard *RGB* video feed, and thus are easily separated into respective color channels for analysis. What follows may be straightforwardly extended to greater or lesser numbers of color channels. We will assume that the observations are geo-referenced to the $(x, y, z) = (0, 0, 0)$ point on the ground plane, which is also assumed to be maintained at camera boresight.

A series of image transformations relate the 3–D scene, $\underline{v}$, to each of K image observations, $\underline{y}_k$, as follows:

$$\underline{y}_k = DBG_kO_k(\underline{v})\underline{v} + \underline{n}_k, \qquad k = 1 \dots K. \tag{2.1}$$

Assume a vectorized 3–D scene volume $\underline{v}$, which is already discretized such that a position on the array $\underline{v}_i, i = \{1 \dots N\}$ uniquely maps to a corresponding (x, y, z) 3–D position of a voxel in $\underline{v}$. The first operation that is applied to this volume for a given observation is the obscuration operator, $O_k(\underline{v})$. This operator is an $N \times N$ diagonal matrix with binary $(0, 1)$ entries that nulls voxels that are not observed at a given aspect. An element of $\underline{v}$ will be nullified by $O_k(\underline{v})$ if it is obscured by other voxels or

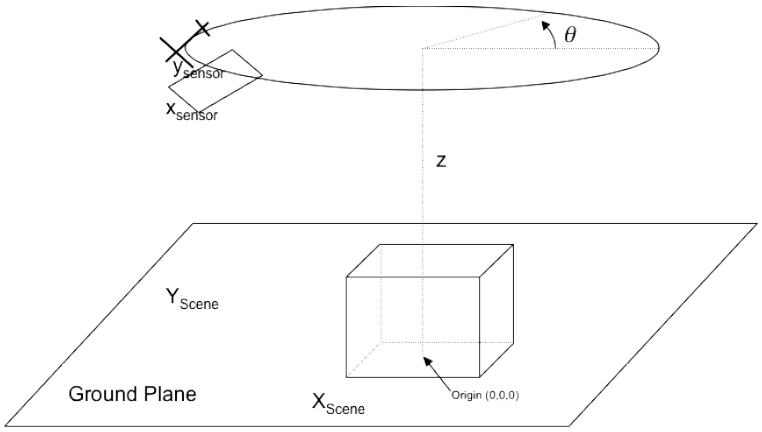

Figure 2.1: Sensor measurement geometry for a point target with position (x, y, z) . Sensors are configured according to their aspect angles in azimuth ϕ and elevation θ.

if it is not filled, in which case that entry in $O_k(\underline{v})$ will remain equal to zero for all aspects k. This operation thus partitions $\underline{v}$ into three types of voxels. The first type includes voxels that are observed as filled for any of the K observations (e.g., a voxel on the surface of a structure in the scene). The second type of voxel is not filled and thus not observed for any of the K observations (e.g., an empty voxel between the observer and a scene structure). The third type of voxel is filled but unobservable for all K observations (e.g., a voxel located in the interior of a closed structure).

Next, G_k performs a geometric transformation (e.g., perspective, orthographic), mapping the 3–D scene coordinates to the 2–D image plane (per observation k). Matrix G_k has dimensions of $M \times N$, where M is the discretized image plane pixel count.

The optical distortion operator, B, and image resampling operator, D, are then applied to the observed 2–D image. The distortion operator is a combination of traditional optical aberrations, such as blurring and warping. The resampling operator downsamples the incoming optical energy and maps the scene image to the sensor-specific detector pixel array. The combination of these two operators dictates how the scene view becomes the captured frame pixel values. For our problem, we will proceed under the assumption that these effects are negligible or are adequately

mitigated by existing methods (i.e., $DB \approx I_{M\times M}$). These operators result in the measured observation, $\underline{y}_k$, a vector of image pixels of size $(M \times 1)$ corrupted by Gaussian measurement noise, $\underline{n}_k$ with $E\{\underline{n}_k \underline{n}_k^T\} = \sigma^2 I$. For a given data collection, an ensemble of such observations are obtained at K nominally distinct aspects.

The resultant observed images are stored, along with the necessary auxiliary information needed for geo-referencing the measurements to the scene coordinates, to be used as input to the tomosynthesis algorithm in order to reconstruct the 3–D scene.

CHAPTER 3
TOMOSYNTHETIC IMAGING

Given a set of 2–D measurements $\{\underline{y}_1 \dots \underline{y}_K\}$, we wish to reconstruct the 3–D volume $\underline{v}$. Popular methods to perform this reconstruction include computed tomography (CT) and tomosynthesis. Tomosynthesis [5] refers to a set of reconstruction techniques that compute a 3–D volumetric image representation from multiple 2–D projections. Tomosynthesis differs from CT in that tomosynthetic volumes are reconstructed from a relatively small number of projections with smaller projection angles. In CT, volume images are reconstructed from many projections and the projection angles are relatively wide.

In traditional tomosynthesis, iterative reconstruction methods are used [12]. These reconstruction techniques utilize some form of the backprojection algorithm to approximate the imaged volume. The basic idea of backprojection is to smear each projection through the reconstructed volume along the same direction the projection was acquired. Therefore, for backprojection to accurately reconstruct a volume, the projection imaging geometry must be accurately known prior to reconstruction.

Some of the assumptions in traditional tomosynthetic imaging are as follows. As mentioned above, imaging algorithms typically neglect the impact of optical distortion, assuming that corrections have already been applied. In addition, image domain sampling is assumed to sufficiently exceed the Nyquist rate such that resampling can be performed efficiently without artifacts. It is typically assumed that the approximation $DB \approx I$ holds. In addition, G_k is assumed known, given acceptable platform stabilization and camera pointing control. Although, compensation of residual in-

ertial errors is still an area of ongoing work. The obscuration operation, $O_k(\underline{v})$, is usually unknown, and thus assumed to be $O_k(\underline{v}) = I \ \forall \ k$. Therefore, the imaging algorithm functions as though all voxels in the scene are filled and visible throughout the data collection, and consequently, all voxels in the resultant tomographic volume are filled. These assumptions transform our measurement model (2.1) into

$$\underline{y}_k = G_k \underline{v} + \underline{n}_k, \tag{3.1}$$

which has the following structure when aggregated over all K measurements:

$$\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_K \end{bmatrix} = \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_K \end{bmatrix} \underline{v} + \begin{bmatrix} \underline{n}_1 \\ \underline{n}_2 \\ \vdots \\ \underline{n}_K \end{bmatrix}. \tag{3.2}$$

The least squares solution for reconstruction of $\hat{\underline{v}}$ is then

$$\hat{\underline{v}} = \left(G^T G\right)^{-1} G^T \underline{y}, \tag{3.3}$$

which is unfortunately computationally prohibitive due to the complexity of the matrix inversion. Instead, tomosynthesis algorithms resort to simply using the adjoint operator $\left(G^T\right)$, resulting in the well-known back-projection algorithm;

$$\hat{\underline{v}} = G^T \underline{y} = \begin{bmatrix} {G_1}^T \dots {G_K}^T \end{bmatrix} \begin{bmatrix} \underline{y}_1 \\ \vdots \\ \underline{y}_K \end{bmatrix} = \sum_{k=1}^{K} {G_k}^T \underline{y}_k \tag{3.4}$$

This approach smears the energy throughout the imaged volume, requiring extensive post-proccessing and other assumptions in order to obtain a useful tomographic volume [4].

CHAPTER 4
EM TOMOSYNTHESIS

As discussed above, the output of traditional tomosynthesis is a volume in which all of the voxels are nonzero. This is because the obscuration operator is assumed to be $O_k(\underline{v}) = I \; \forall \; k$. Our novel solution to the tomosynthesis problem is to use the adjoint operator, but to iteratively estimate $O_k(\underline{v})$ through the expectation-maximization algorithm. This will eliminate many of the filled voxels that do not, in fact, represent the original scene volume. The tomosynthetic measurement equation (3.2) thus becomes

$$\underline{y}_k = G_k O_k(\underline{v})\underline{v} + \underline{n}_k \tag{4.1}$$

which has the following structure when aggregated over all measurements K:

$$\underline{y} = \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_K \end{bmatrix} = \begin{bmatrix} G_1 O_1(\underline{v}) \\ G_2 O_2(\underline{v}) \\ \vdots \\ G_K O_K(\underline{v}) \end{bmatrix} \underline{v} + \begin{bmatrix} \underline{n}_1 \\ \underline{n}_2 \\ \vdots \\ \underline{n}_K \end{bmatrix}. \tag{4.2}$$

The tomosynthetic volume estimate is then given by

$$\hat{\underline{v}} = \sum_{k=1}^{K} \hat{O}_k^T(\underline{v}) {G_k}^T \underline{y}_k, \tag{4.3}$$

where, as will be detailed in Chapter 5, $\hat{O}_k(\underline{v})$ is estimated in the expectation step and (4.3) represents calculation of the maximization step. Iterative refinement of the

obscuration operator ensures that energy is back-projected only into the voxels of $\underline{v}$ that were observed by the sensor system. This greatly mitigates the incorrect smearing of energy through both filled and unfilled voxels. The proposed EM approach will thereby produce a volume estimate that will not only have a more accurate spatial representation of the true volume $\underline{v}$, but will also have a more accurate spectral (RGB) color representation. This is directly extensible to higher-dimensioned color spaces (e.g., hyper-spectral imaging).

CHAPTER 5
ALGORITHM IMPLEMENTATION

A common need in sensor data processing is the estimation of the parameters of a probability distribution function based on a finite set of observations. In cases where multiple distributions are to be estimated and observations are randomly drawn from those distributions, association of observations to distributions is necessary for reliable parameter estimation. For the well-known data clustering problem, the expectation-maximization (EM) algorithm is commonly used to classify data observations and estimate the parameters of their underlying distributions. There, a set of observations

$$\underline{y} = \{(c_1, x_1) \dots (c_L, x_L)\} \tag{5.1}$$

are randomly drawn from a set of distributions

$$P = \{p\left(X|\Theta_c\right)|c = 1 \dots C\} \tag{5.2}$$

where the distribution class label c_l for each observation is unknown. Application of the EM algorithm is an iterative two-step process. In the expectation step, observations are assigned to classes via $\hat{c}_l = E\{c_l|\underline{x}_l\}$. The distribution parameters are then updated in the maximization step via the maximum likelihood estimate

$$\underline{\hat{\Theta}}_c = \arg\max_{\underline{\Theta}} p\left(\underline{y}_c|\underline{\Theta}\right) \tag{5.3}$$

where $\underline{y}_c$ is the subset of observations assigned to class c.

The above proposed tomosynthesis approach may be formulated as an EM algorithm as follows. Over the course of its data collection trajectory the sensor aggregates the set of observations $\{(c_1, \underline{x}_1) \dots (c_L, \underline{x}_L)\}$ where $L = KM$, K frames of M pixels each. The $\underline{x}_l$ are RGB vectors from $\underline{y}$ in (4.2) representing the recorded pixel values, and the unknown c_l indicate the observation distribution from which each was drawn. In this case, the c_l correspond to the indices of voxels within the imaged volume. Given additive white Gaussian measurement noise in (4.2), we assume that each $\underline{x}_l$ represents an observation of a single voxel RGB vector $\underline{v}_{c_l}$ such that $\underline{x}_l$ is vector Gaussian with mean equal to the observed voxel's RGB values and with covariance $\Sigma = \sigma^2 I$. Therefore, in the E step, we seek to assign each pixel observation $\underline{x}_l$ to an image voxel $\underline{v}_{\hat{c}_l}$ where $\hat{c}_l = E\{c_l | \underline{x}_l\}$. This selection is limited to the subset of voxels that fall on the line-of-sight of the l_{th} pixel observation, as specified by the geometric operator G_k for a given aspect. We designate this voxel subset as V_l. This classification is performed via likelihood ratio test such that

$$\hat{c}_l = \arg\max_{c_l \in V_l} p\left(\underline{x}_l | \hat{\underline{\Theta}}_{c_l}\right) = \arg\max_{c_l \in V_l} \left(\frac{1}{(2\pi)^{\frac{3}{2}} \left|\hat{\Sigma}_{c_l}\right|^{\frac{1}{2}}} e^{-\frac{1}{2}\left(\underline{x}_l - \hat{\underline{v}}_{c_l}\right)^T \hat{\Sigma}_{c_l}^{-1} \left(\underline{x}_l - \hat{\underline{v}}_{c_l}\right)} \right) \tag{5.4}$$

where the parameter vector $\hat{\underline{\Theta}}_{c_l}$ represents the current estimate of the mean ($\hat{\underline{v}}_{c_l}$) and covariance ($\hat{\Sigma}_{c_l}$) of voxel $\underline{v}_{c_l}$. The maximization step then updates the voxel parameter estimates via the maximum likelihood estimate

$$\hat{\underline{\Theta}}_{c_l} = \arg\max_{\underline{\Theta}} p\left(\underline{y}_c | \underline{\Theta}\right) \tag{5.5}$$

$$\hat{\underline{\Theta}}_c \rightarrow \begin{array}{ll} \hat{\underline{v}}_c = & \left\langle \underline{y}_c \right\rangle \\ \hat{\Sigma}_c = & \left\langle \left(\underline{y}_c - \hat{\underline{v}}_c\right)\left(\underline{y}_c - \hat{\underline{v}}_c\right)^T \right\rangle \end{array}$$

where $\underline{y}_c$ represents the subset of pixel observations in $\underline{y}$ that have been classified as

members of cluster c, meaning observations of pixel $\underline{v}_c$. In (5.5), $\langle\cdot\rangle$ represents the arithmetic mean.

The obscuration operator $\hat{O}_k(\underline{v})$ is estimated in the E step through the classification process implemented via (5.4). By assigning each pixel observation to a single image voxel we are effectively designating the unobscured voxels in $O_k(\underline{v})$ for each aspect. Figure 5.1 illustrates the evolution of both $\hat{c}_l$ and $\hat{O}_k(\underline{v})$. The left side of the figure shows the output of a traditional tomosynthesis volume (top left) and the traditional $O_k(\underline{v}) = I$ (bottom left). The likelihood calculation is performed to satisfy (5.4) (center), and the resultant voxel estimate of the volume $\hat{c}_{l_k}$ (top right) and $\hat{O}_k(\underline{v})$ (bottom right).

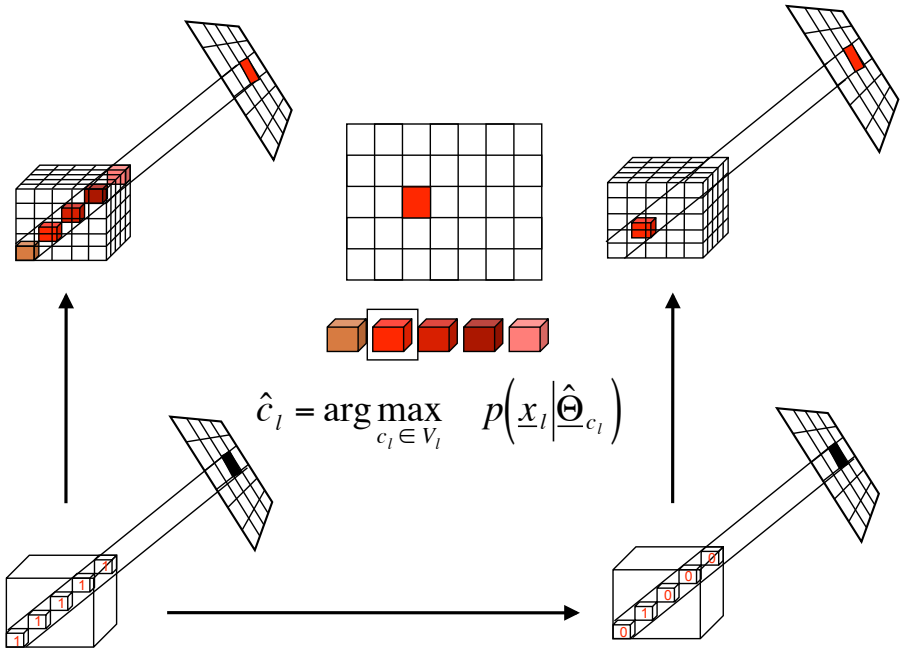

Figure 5.1: Pictorial representation of the evolution of $\hat{c}_l$ and $\hat{O}_k(\underline{v})$ for a single EM iteration.

In practice, this process may be implemented quite efficiently. Given sufficient memory to maintain new and old estimates of the parameters $\underline{\Theta} = \{\underline{\Theta}_1 \ldots \underline{\Theta}_N\}$, one may iterate the tomosynthesis imaging process, updating class estimates $\hat{c}_l$ and parameter estimates simultaneously. For a given pixel observation, the adjoint operator G_K^T geometrically restricts the number of voxels that must be considered in the E step,

and the linearity of the Gaussian measurement model allows parameter estimates to be updated sequentially for the M step. To prevent early classifications from leading to an unrecoverable state in the EM process, we initially allow pixel observations to be assigned to multiple geometrically-acceptable voxels and reduce that number as parameter estimates converge. We also place a minimum threshold on the number of pixel observation assignments that a voxel must receive in order for it to remain "active". Once a voxel has been deactivated, it is removed from the assignment classifications and is not displayed in the final image. This culling of infeasible voxels from further computations can obviously produce dramatic computational savings.

CHAPTER 6

SIMULATION DESCRIPTION

6.1 Scene Generation

We implement a scene consisting of 4 colored walls placed atop a structured ground plane. This is generated by defining a 3–D RGB point model in an xyz volume matrix such that every populated (x, y, z) element of the scene volume has corresponding values for RGB. Figure 6.1 shows this scene, which was implemented in MATLAB by generating planes of various colors and assigning each voxel in the volume an x, y, z value for position, while concurrently assigning RGB values for the red, green, and blue components.

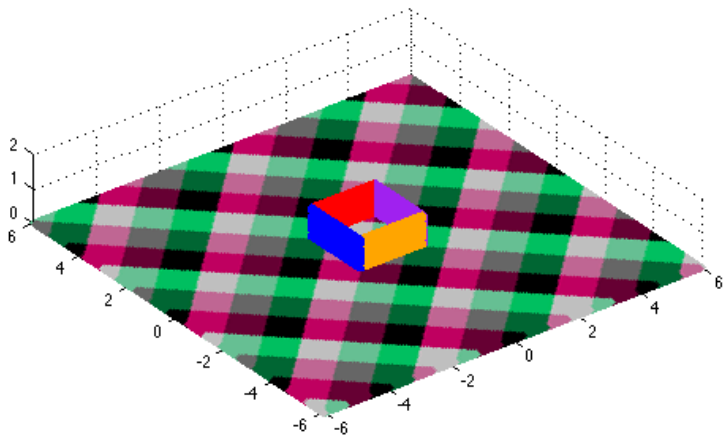

Figure 6.1: Simulated scene of 4 walls constructed on a multi-colored ground plane.

6.2 Scene Capture

The imaging sensor is simulated as a simple video imager capturing independent RGB values onto an 81×81 image plane, giving us $M = (81 \times 81) = 6,561$. We implement equation (3.2) by "flying" the sensor around the scene, capturing an image ($\underline{y}_k$) every 10° in azimuth, giving us $k = 1 \ldots K$ observations with $K = 36$, and $\theta = k\,[i] = 10, 20 \ldots 360$. This gives us $L = KM = (36 \times 6,561) = 236,196$. For this simulation we kept a constant elevation angle of $\phi = 25^\circ$.

G_k operates on the volume $\underline{v}$ to project the indices of each voxel onto the geometrically derived position on the imager array $\underline{y}$. For this problem we assume an orthographic projection geometry. The mapping of a 3–D voxel's coordinates (x_v, y_v, z_v) onto a 2–D image pixel location (x_p, y_p) is thus computed as

$$\begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} -\sin\theta_k & \cos\theta_k & 0 \\ -\cos\theta_k \sin\phi_k & -\sin\theta_k \sin\phi_k & \cos\theta_k \end{bmatrix} \begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix}. \tag{6.1}$$

Gaussian noise ($\underline{n}_k$) with standard deviation σ is added to each of the RGB image channels independently during the scene capture. Figure 6.2 shows the RGB composite scene as well as the individual RGB image planes for an observation with a noise level of $\sigma = 0.05$ captured at an azimuth of $\theta = 40^\circ$ and an elevation of $\phi = 25^\circ$. The image plane's (x, y) pixels are represented by the coordinates. Each image plane will return a color intensity $RGB \in [0, 1]$.

6.3 Classical Tomosynthesis

Now that we have the set of $K = 36$ observations, a classical tomosynthesis is performed. We determine the projection vectors for each of the image pixels into

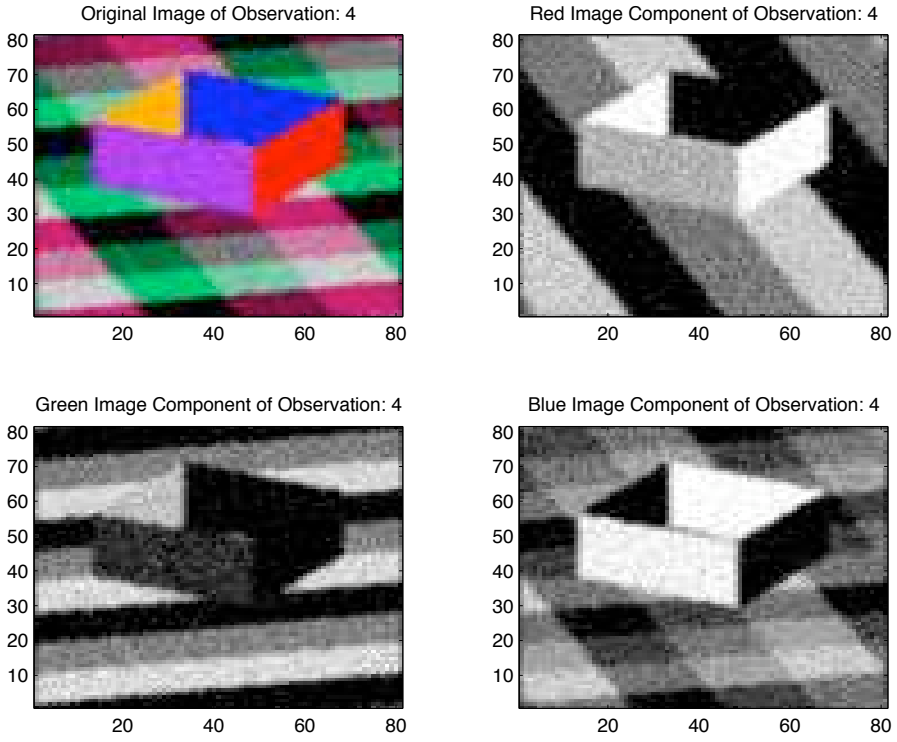

Figure 6.2: Simulated observation of the 4 wall scene captured at $\theta = 40°$ and $\phi = 25°$ with a noise level of $\sigma = 0.05$.

the tomosynthetic volume using the G_k from the scene capture equation (3.2) to implement the back-projection equation (3.4) and then project the RGB intensities throughout the volume along that projection vector. The tomosynthetic volume then accumulates the energy of the K observations to complete the creation of the volume. The second-order covariance matrices needed for the likelihood calculations in equation (5.4) are also formed here. The output of the tomosynthesis is shown in Figure 6.3. Generally, all voxels are filled and thresholding is used to display the voxels with the highest accumulated backprojected energy.

6.4 Expectation–Maximization Algorithm

At this point, we have all the information needed to implement the expectation–maximization algorithm: the original K observations $\underline{y}_k$, the classical tomosynthesis

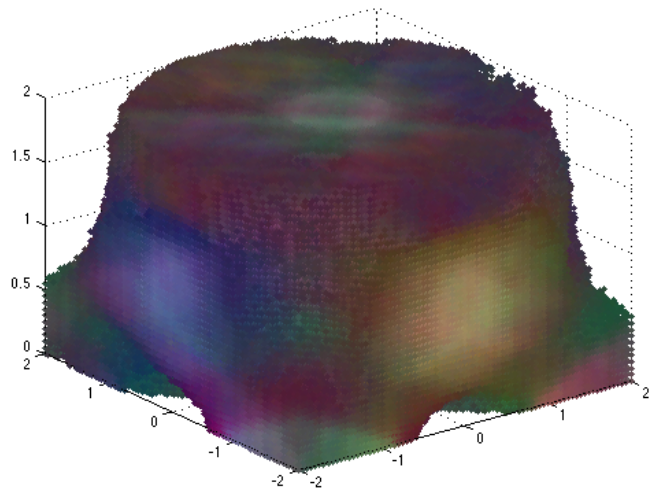

Figure 6.3: Output of the classical tomosynthesis module.

output volume (which serves as the initial estimate $\underline{\hat{v}}$), and the covariance matrices for the RGB values. We choose the initial estimate of $\hat{O}_k(\underline{v})$ to be I. Iteration over all K observations is conducted to iteratively assign pixel observations to voxels (E-step) and estimate the voxel parameters (M-step).

Using our known G_k, we determine which of the voxels in the input tomosynthetic volume are along the projection vector of each image pixel. For the initial iteration of the EM algorithm, we process every voxel corresponding to the image pixel projection ($\hat{O}_k(\underline{v}) = I$). We then iterate over each of the voxels along $\hat{O}_k(\underline{v})$ and determine the likelihood value for each using equation (5.4). We set a contribution limit, Γ_i for each iteration of the EM algorithm, which controls the convergence rate of the algorithm. For the presented simulation $\Gamma = \{7, 5, 3, 1\}$. We then select the Γ_i voxels with the highest likelihood values and implement equation (5.5) by populating the corresponding voxels in an identically-sized accumulator volume with the RGB values of the image pixel. In addition *for each scene voxel*, we track the number of times that a voxel received a contribution in the current iteration and assign this number, which tracks the number of image pixels assigned to each scene voxel, to $\underline{\alpha}$. Voxels whose value in $\underline{\alpha}$ falls below α_{min} will be deactivated, meaning they will be eliminated

from further EM iterations and will not be displayed in the final image.

We will illustrate this with intermediate data products for $k = 4$ ($\theta = 40°$). Since we iterate over every image pixel, we chose a pixel of interest (POI) at $(x, y) = (30, 40)$. Figure 6.4 shows the POI as a white dot for clarity. Note that MATLAB's internal representation of pixel coordinate $(x, y) = (30, 40)$ corresponds to a linear pixel number of 2389.

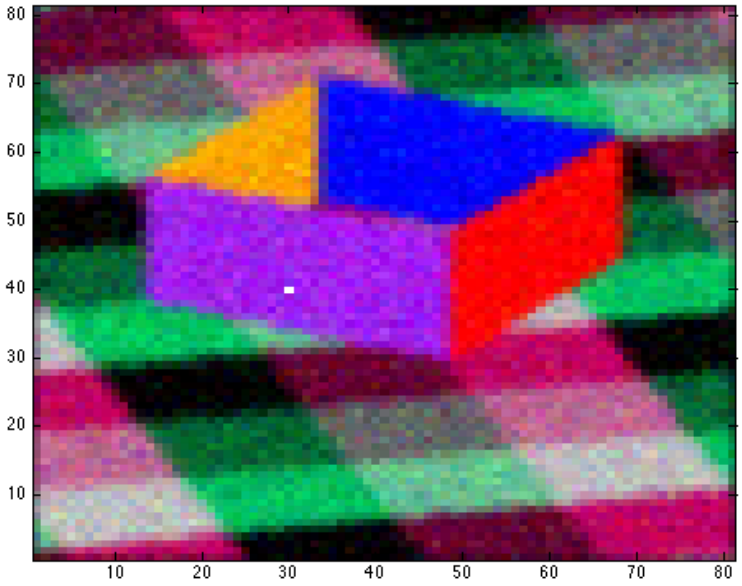

Figure 6.4: Pixel of interest for example iteration of the EM step.

Figure 6.5 shows the RGB values of geometrically acceptable voxels, indexed in arbitrary order, falling along the back-projection from the POI as specified by G_k^T. Figure 6.6 displays the output from the likelihood calculation for those same voxels.

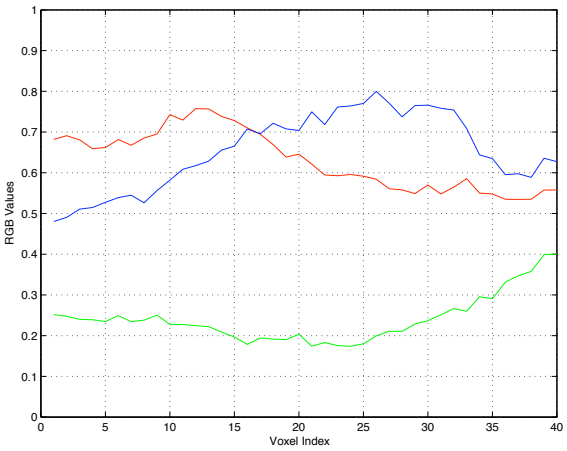

Figure 6.5: *RGB* values of the tomosynthesis volume used in the likelihood calculation

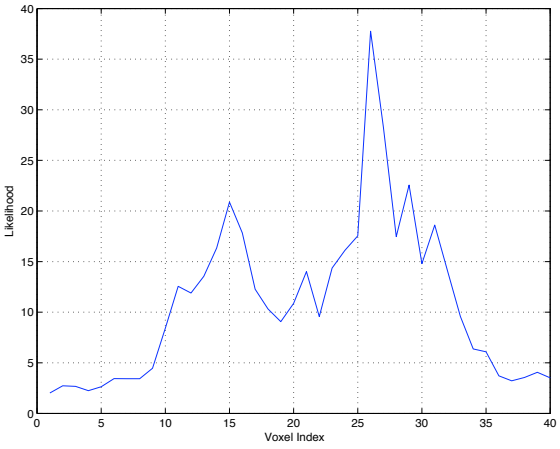

Figure 6.6: Likelihood values for the voxels corresponding to the POI

Figure 6.7 shows the voxels corresponding to the projection of the image pixel though the volume with a marker on the voxel corresponding to the maximum likelihood selection. Figure 6.8 shows the voxels corresponding to the POI from a different observation angle. As the EM algorithm progresses, the selected voxels will converge to match those selected by the obscuration operator ($\hat{O}_k(\underline{v}) \rightarrow O_k(\underline{v})$).

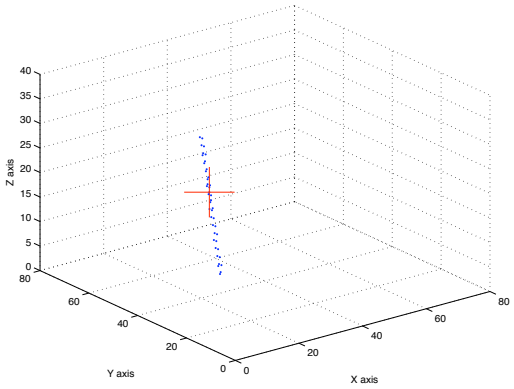

Figure 6.7: Voxels corresponding to the POI 2389 with Maximum-Likelihood selection indicated.

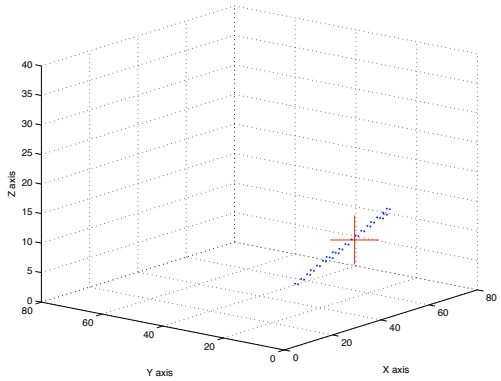

Figure 6.8: Voxels corresponding to the POI 2389 with Maximum-Likelihood selection indicated from an observation 180° from Figure 6.7.

The next iteration of the EM algorithm takes the resultant accumulator volume derived in the previous iteration and performs the likelihood calculation on the volume of RGB values, this time accumulating a storage volume for $\Gamma_i = 5$. This is iterated over $\Gamma_i = 3$ and then finally $\Gamma_i = 1$. The volume corresponding to the output of the intermediate iteration ($\Gamma_i = 5$) is shown in Figure 6.9. The volume corresponding to the output of the final iteration ($\Gamma_i = 1$) is shown in Figure 6.10. For display purposes, we chose $\alpha_{min} = 4$. Note the intermediate EM tomosynthesis output (Figure 6.9) shows a vast improvement over the output of the classical tomosynthesis (Figure 6.3). We see structure forming, with the voxels coalescing towards a cohesive spatial estimate. The final output of our algorithm shows a fairly sharply-defined shape with some false structure voxels that we will quantitatively discuss later.

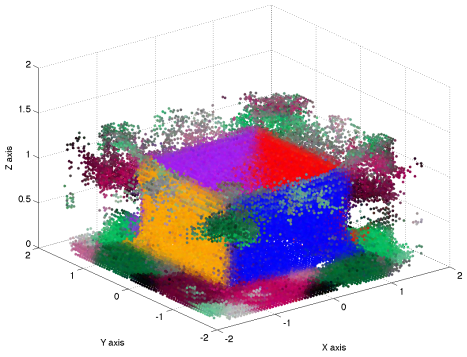

Figure 6.9: Output volume of EM iteration containing accumulated voxels at $\Gamma_i = 5$ and $\alpha_{min} = 4$ for the 4-wall scenario.

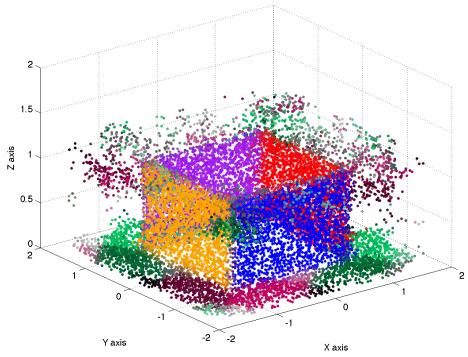

Figure 6.10: Output volume of final EM iteration for the 4-wall scenario ($\alpha_{min} = 4$).

We generated other scenes to test the algorithm over different conditions. An example is a scene where an image of the venerable Tweety Bird, shown in Figure 6.11, is placed on top of the 4 walls of the previous example. We see the image start to emerge from the classical tomosynthesis volume as seen in the intermediate ($\Gamma_i = 5$) volume shown in Figure 6.12. The final output ($\Gamma_i = 1$) is shown in Figure 6.13.

Figure 6.11: (81×81) pixel image of Tweety Bird used for the Tweety Bird scenario.

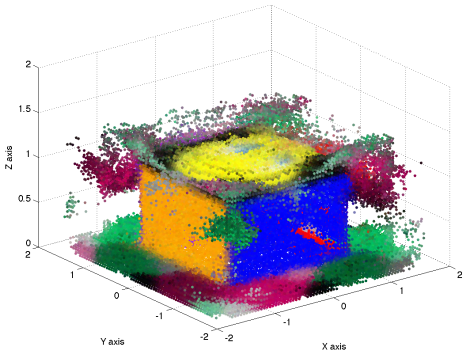

Figure 6.12: Output volume of EM iteration containing accumulated voxels at $\Gamma_i = 5$ and $\alpha_{min} = 4$ for the Tweety Bird scenario.

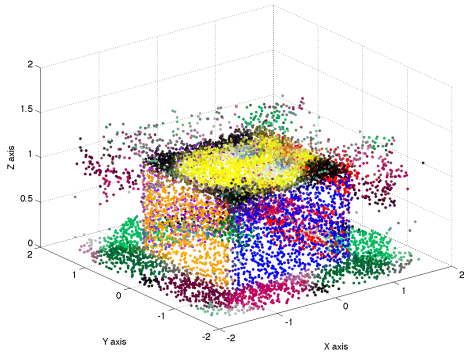

Figure 6.13: Output volume of final EM iteration the Tweety Bird scenario ($\alpha_{min} = 4$).

Another example is a scene where we have only 2 intersecting perpendicular walls. Again, the intermediate volume corresponding to the output of iteration ($\Gamma_i = 5$) is shown in Figure 6.14 and the final output is shown in Figure 6.15. Again, we see the rapid spatial and spectral convergence of the algorithm. For the final EM

tomosynthesis output of the Tweety Bird scenario, we observe that the number of wall voxels estimated by the algorithm is slightly fewer than the corresponding walls of the 4-wall scenario. We attribute this to the number of observations that were available to the Tweety Bird scenario versus the 4-wall scenario. Each of the walls for the Tweety Bird scene were seen by, at most, half of the observations, whereas the 4-wall scene had at least a portion of each wall visible for almost all of the observations.

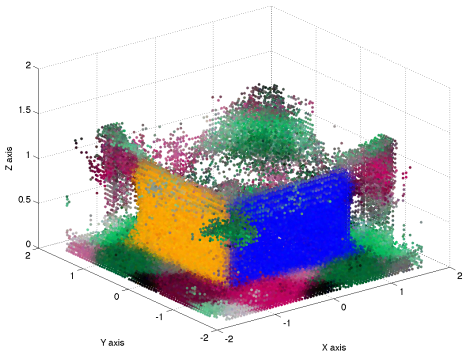

Figure 6.14: Output volume of EM iteration containing accumulated voxels at $\Gamma_i = 5$ and $\alpha_{min} = 4$ for the 2-wall scenario.

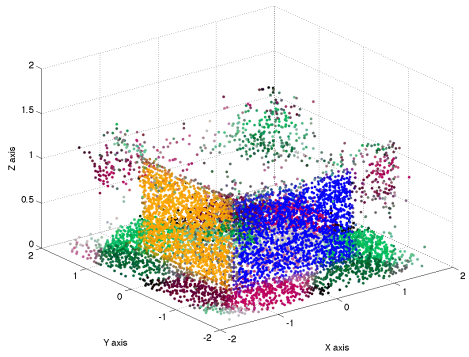

Figure 6.15: Output volume of final EM iteration for the 2-wall scenario ($\alpha_{min} = 4$).

We wish to illustrate the effect of increasing α_{min}, which will decrease the number of voxels being declared (displayed). Taking the previous scene, we increase α_{min} from 4 to 7. We see that the density of the volume decreases, which not only decreases the number of structure voxels, but also decreases the the amount of "noise" voxels floating about the walls. As before, the intermediate volume corresponding to the output of iteration ($\Gamma_i = 5$) is shown in Figure 6.16 and the final output is shown in Figure 6.17.

All of the previous examples were multispectral scenes constructed of flat surfaces. We now examine results that test the algorithm under monochromatic observations as well as observations of curved surface structures. The EM tomosynthesis volume for a four-walled scene simulated as a grayscale observation is shown in Figure 6.18.

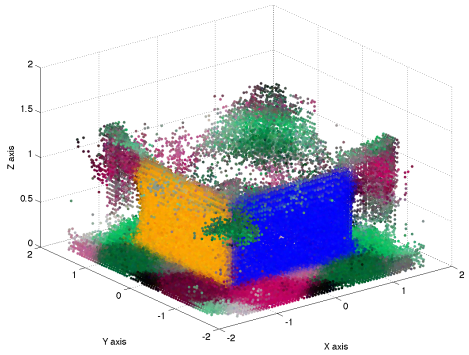

Figure 6.16: Output volume of EM iteration containing accumulated voxels at $\Gamma_i = 5$ for the 2-wall scenario with $\alpha_{min} = 7$.

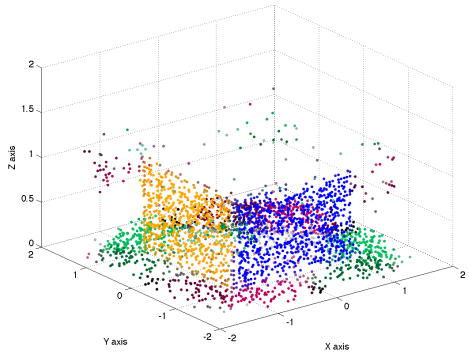

Figure 6.17: Output volume of final EM iteration for the 2-wall scenario with $\alpha_{min} = 7$.

This time there is considerably more noise voxels than the results shown in Figure 6.10. This is due to the likelihood function having reduced discriminatory capability without color information.

Figure 6.19 demonstrates the EM tomosynthetic output of an observed scene constructed of convex and concave multispectral walls. These two walls replace the flat blue and purple walls shown in Figure 6.10. A "roof" constructed of the same flat image of Tweety Bird as shown in Figure 6.13 is added atop the four walls. Comparing this result to Figure 6.13 we see that the noise voxels appear to be on the same order as the flat-walled counterpart, and it appears that the EM tomosynthesis algorithm is capable of reconstructing both convex and concave surfaces.

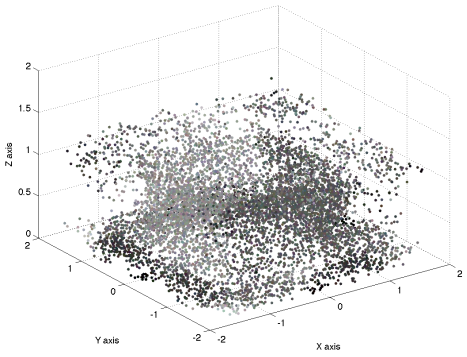

Figure 6.18: Final output volume of EM tomosynthesis containing accumulated voxels for a grayscale observation of the 4-wall scenario with $\alpha_{min} = 5$.

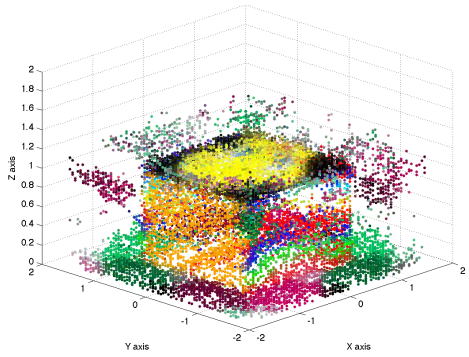

Figure 6.19: Final output volume of EM tomosynthesis containing accumulated voxels for a curved surface (with "roof") scenario with $\alpha_{min} = 4$.

CHAPTER 7
PERFORMANCE ANALYSIS

In order to characterize performance, a "truth volume" with a voxel spacing equivalent to the tomosynthesis output volume was generated. This is shown in Figure 7.1.

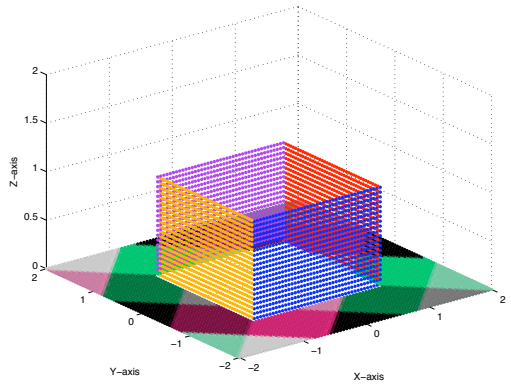

Figure 7.1: Truth volume for 4-wall scenario.

Our algorithm estimates both spatial and spectral properties of the scene. To measure performance of the spatial estimation, we determine how many of the voxels exceeding the likelihood selection count α_{min} matched the location of the truth volume voxels within a 3-voxel membrane around the truth voxel locations. The volume of correct declarations is shown in Figure 7.2 for an EM output volume of voxels having a likelihood count $\alpha_{min} = 1$ (meaning every voxel with at least 1 EM contribution). The volume of voxels which were part of the truth volume that were not declared by the EM algorithm (i.e., missed detects) is shown in Figure 7.3.

Finally, the voxels that were declared as detects that were, in fact, not part of the

scene is shown in Figure 7.4. For this figure, we scaled the false detect output by the strength of the voxel's value of $\underline{\alpha}$. An interesting result is that even though there are a number of false detects, the likelihood associated to the vast majority of those false detects is quite low.

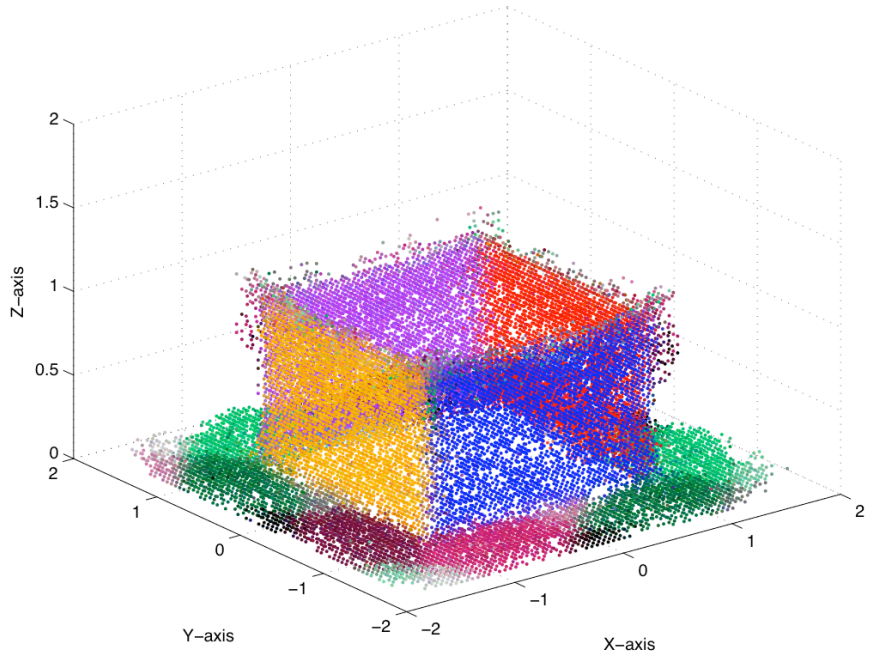

Figure 7.2: Correct declaration of the EM algorithm for likelihood count $\alpha_{min} = 1$.

The volume of correct declarations for the convex scene introduced in Figure 6.19 is shown in Figure 7.5, illustrating the capability of the EM algorithm to estimate concave and convex surfaces. Figure 7.6 shows the truth volume for this scene.

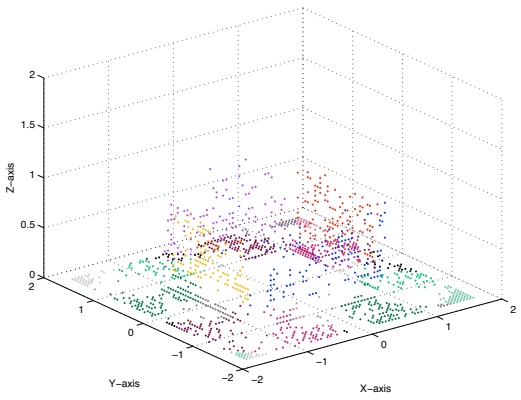

Figure 7.3: Missed detects of the EM algorithm for likelihood count $\alpha_{min} = 1$.

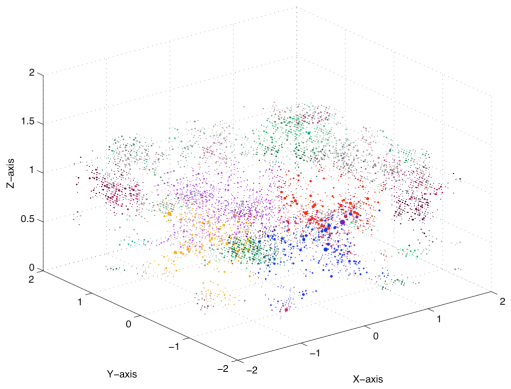

Figure 7.4: False detects of the EM algorithm for likelihood count $\alpha_{min} = 1$.

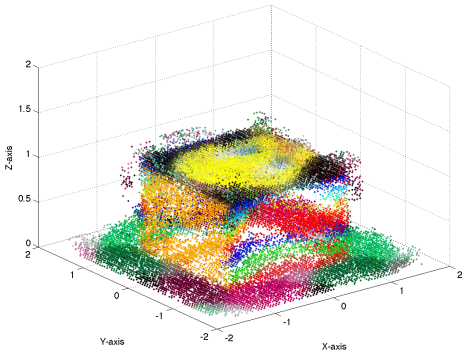

Figure 7.5: Correct declaration of EM tomosynthesis for convex scenario with $\alpha_{min} = 1$.

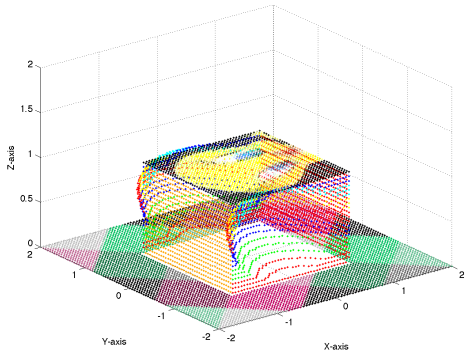

Figure 7.6: Truth volume for convex scenario.

We formally characterize the performance of the algorithm with a parametric excursion over α_{min} for a range of noise levels ($\sigma = \{0.05, 0.15, 0.25\}$) to obtain a traditional receiver operating characteristic (ROC) curve. This characterization is presented for the 4-wall scenario, and is shown in Figure 7.7 for the 25° elevation.

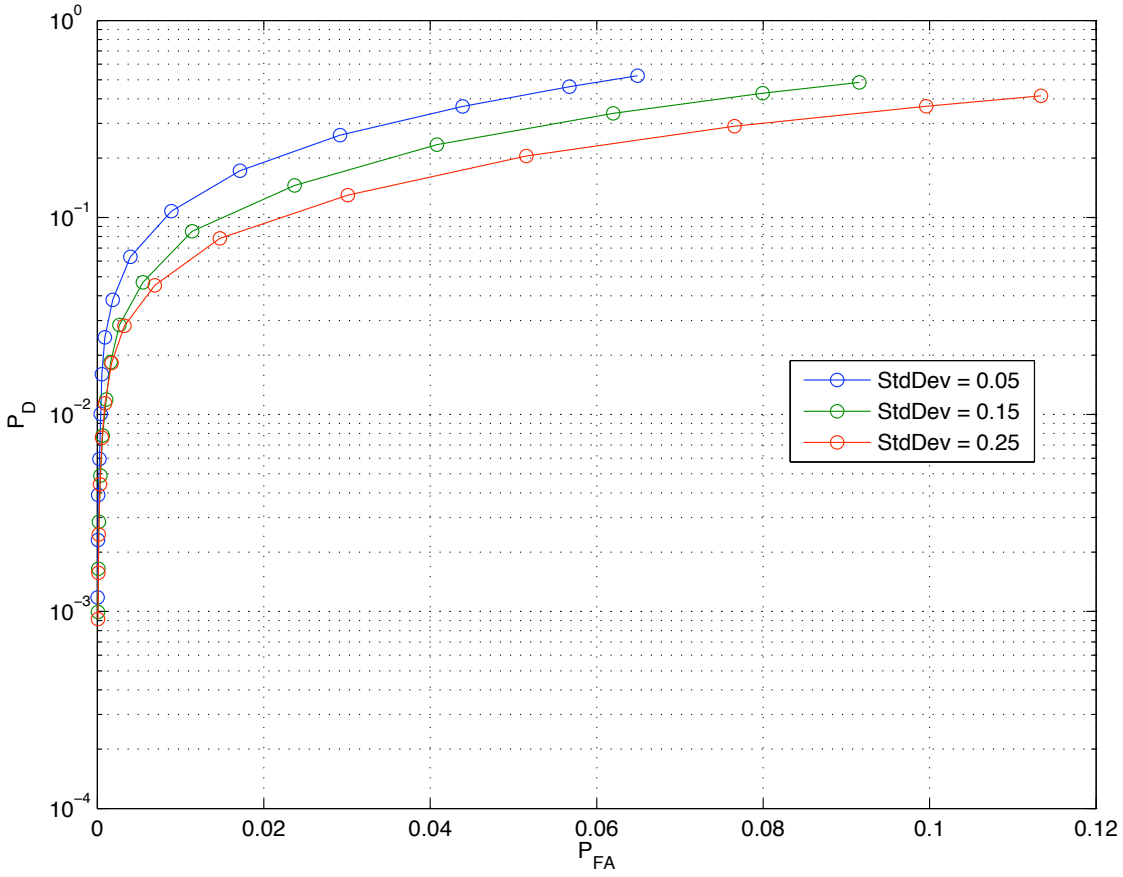

Figure 7.7: General receiver operating characteristic curve for 25° elevation.

In addition to evaluating the spatial accuracy of the algorithm, we also wish to evaluate the accuracy of the EM color estimates. A histogram of the error between the true spectral characteristics of the scene volume and the estimated spectral characteristics of the EM algorithm is presented in Figure (7.8). The MSE was determined by taking the mean value of the square of the differences between the RGB values of the true scene and the estimated scene on a voxel by voxel basis, which for this instance of $\alpha_{min} = 1$ was only 0.0073

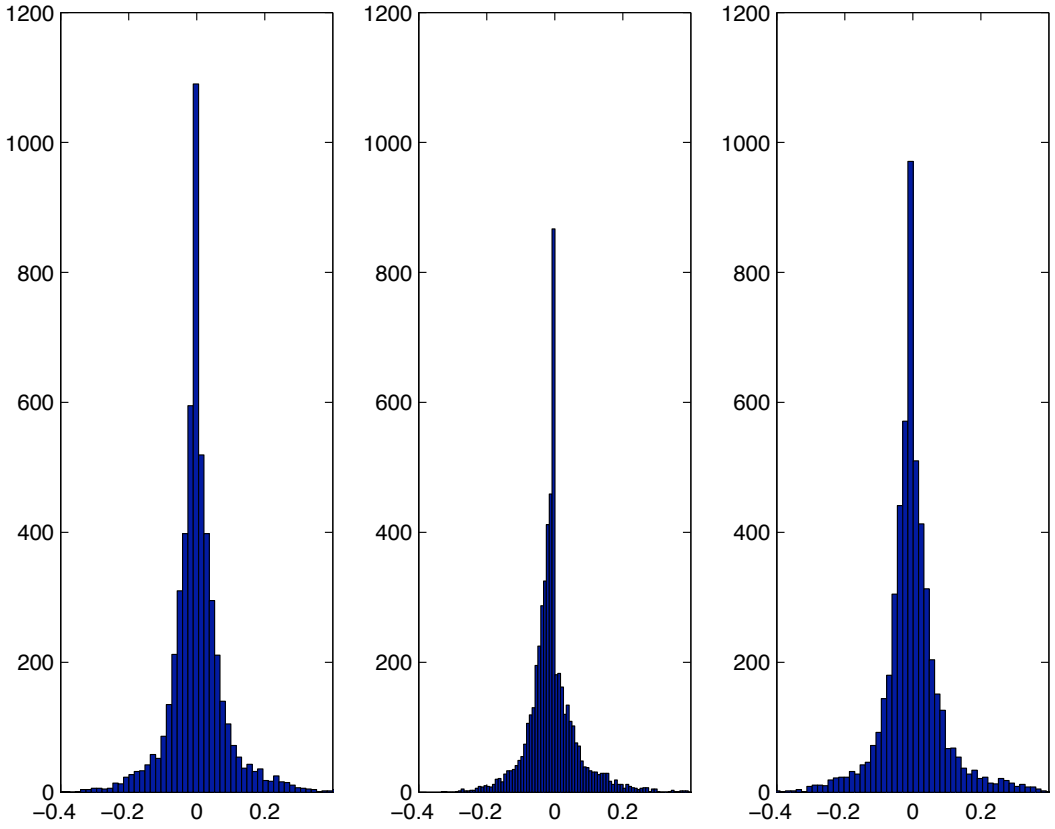

Figure 7.8: Red (left panel), Green (middle panel), and Blue(left panel) color error between the true scene and the EM estimate of RGB value for 25° elevation and $\sigma = 0.05$, $\alpha_{min} = 1$.

CHAPTER 8
CONCLUSION AND FUTURE WORK

We have presented an application of the expectation-maximization estimator applied to tomosynthesis of image sequences. This approach shows the ability to work in a high-noise observation environment, as indicated by the ROC plots and the ability to estimate the spectral characteristics of the scene, with an extremely small MSE. In particular, it would be an interesting addition to incorporate a non-binary $O_k(\underline{v})$ to the observation/reconstruction model, such that the elements of $O_k(\underline{v}) \rightarrow 0 \leq \beta \leq 1$, thus allowing for varying opaqueness of structures.

REFERENCES

[1] Pei Chen and K. Barner, *Three-dimensional multi-resolution statistical reconstruction for tomosynthesis*, Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on, (15-18 April 2004), pp. 559–562 Vol. 1.

[2] ——, *Maximum likelihood reconstruction for tomosynthesis*, Bioengineering Conference, 2003 IEEE 29th Annual, Proceedings of, (22-23 March 2003), pp. 59–60.

[3] A.R. De Pierro and M.E.B. Yamagishi, *Fast EM-like methods for maximum "a posteriori" estimates in emission tomography*, Medical Imaging, IEEE Transactions on, 20 (Apr 2001), pp. 280–288.

[4] D.T. Gering and III Wells, W.M., *Object modeling using tomography and photography*, Multi-View Modeling and Analysis of Visual Scenes, 1999. (MVIEW '99) Proceedings. IEEE Workshop on, (1999), pp. 11–18.

[5] David G. Grant, *Tomosynthesis: A three-dimensional radiographic imaging technique*, Biomedical Engineering, IEEE Transactions on, BME-19 (Jan. 1972), pp. 20–28.

[6] R.M. Lewitt and S. Matej, *Overview of methods for image reconstruction from projections in emission computed tomography*, Proceedings of the IEEE, 91 (Oct. 2003), pp. 1588–1611.

[7] J. Li and R. Chellappa, *Structure from planar motion*, Image Processing, IEEE Transactions on, 15 (Nov. 2006), pp. 3466–3477.

[8] B. Lipinski, H. Herzog, E. Rota Kops, W. Oberschelp, and H.W. Muller-Gartner, *Expectation maximization reconstruction of positron emission tomography images using anatomical magnetic resonance information*, Medical Imaging, IEEE Transactions on, 16 (Apr 1997), pp. 129–136.

[9] J.M. Ollinger, *Iterative reconstruction-reprojection and the expectation-maximization algorithm*, Medical Imaging, IEEE Transactions on, 9 (Mar 1990), pp. 94–98.

[10] ——, *Maximum-likelihood reconstruction of transmission images in emission computed tomography via the EM algorithm*, Medical Imaging, IEEE Transactions on, 13 (Mar 1994), pp. 89–101.

[11] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang, *Super-resolution image reconstruction: a technical overview*, Signal Processing Magazine, IEEE, 20 (May 2003), pp. 21–36.

[12] T.M. Persons, P.F. Hemler, and R.J. Plemmons, *3D iterative restoration of tomosynthetic images*. Unpublished, 2000.

[13] K.S. Pradeep and A.N. Rajagopalan, *Improving shape from focus using defocus information*, Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, 1 (2006), pp. 731–734.

[14] Y.Y. Schechner and N. Kiryati, *Depth from defocus vs. stereo: how different really are they?*, Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on, 2 (16-20 Aug 1998), pp. 1784–1786 vol.2.

[15] L. A. Shepp and Y. Vardi, *Maximum likelihood reconstruction for emission tomography*, Medical Imaging, IEEE Transactions on, 1 (Oct. 1982), pp. 113–122.

[16] H. Sunyoto, W. van der Mark, and D.M. Gavrila, *A comparative study of fast dense stereo vision algorithms*, Intelligent Vehicles Symposium, 2004 IEEE, (14-17 June 2004), pp. 319–324.

[17] Li Zhang, B. Curless, and S.M. Seitz, *Spacetime stereo: shape recovery for dynamic scenes*, Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, 2 (18-20 June 2003), pp. II–367–74 vol.2.